Received: 4 June 2020 | Accepted: 24 August 2020

DOI: 10.1111/apt.16086

AP&T Alimentary Pharmacology & Therapeutics WILEY

Multiomic features associated with mucosal healing and inflammation in paediatric Crohn's disease

Henry Taylor[1] | Jose Ivan Serrano-Contreras[2] | Julie A. K. McDonald[3] | Jenny Epstein[4] | JM Fell[4] | Rocio C. Seoane[5] | Jia V. Li[2] | Julian R. Marchesi[5,6] | Ailsa L. Hart[2,7]

[1]Department of Surgery and Cancer, Imperial College London, London, UK

[2]Department of Metabolism, Digestion and Reproduction, Faculty of Medicine, Imperial College London, London, UK

[3]MRC Centre for Molecular Bacteriology and Infection, Imperial College London, London, UK

[4]Paediatric Gastroenterology Department, Chelsea and Westminster Hospital, London, UK

[5]Division of Digestive Diseases, Department of Metabolism, Digestion and Reproduction, Imperial College London, London, UK

[6]School of Biosciences, University of Cardiff, Cardiff, UK

[7]IBD Unit, St. Mark's Hospital, Harrow, UK

Correspondence
Ailsa L. Hart, Department of Metabolism, Digestion and Reproduction, Faculty of Medicine, Imperial College London, London, UK.
Email: ailsa.hart@nhs.net

Funding information
This work was funded by the National Institute of Health Research (NIHR) and the NIHR Imperial Biomedical Research Centre (BRC). The Division of Digestive Disease at Imperial College London receives financial support from the National Institute of Health Research (NIHR) Imperial Biomedical Research Centre (BRC) based at Imperial College Healthcare NHS Trust and Imperial College London. This article is independent research funded by the NIHR BRC, and the views expressed in this publication are those of the authors and not necessarily those of the NHS, NIHR or the Department of Health.

Summary

Background: The gastrointestinal microbiota has an important role in mucosal immune homoeostasis and may contribute to maintaining mucosal healing in Crohn's disease (CD).

Aim: To identify changes in the microbiota, metabolome and protease activity associated with mucosal healing in established paediatric CD.

Methods: Twenty-five participants aged 3-18 years with CD, disease duration of over 6 months, and maintenance treatment with biological therapy were recruited. They were divided into a low calprotectin group (faecal calprotectin <100 μg/g, "mucosal healing," n = 11), and a high calprotectin group (faecal calprotectin >100 μg/g, "mucosal inflammation," n = 11). 16S gene-based metataxonomics, ^{1}H-NMR spectroscopy-based metabolic profiling and protease activity assays were performed on stool samples.

Results: Relative abundance of *Dialister* species was six times greater in the low calprotectin group (q = 0.00999). Alpha and beta diversity, total protease activity and inferred metagenomic profiles did not differ between groups. Pentanoate (valerate) and lysine were principal discriminators in a machine-learning model which differentiated high and low calprotectin samples using NMR spectra (R^2 0.87, Q^2 0.41). Mean relative concentration of pentanoate was 1.35-times greater in the low calprotectin group (95% CI 1.03-1.68, P = 0.036) and was positively correlated with *Dialister*. Mean relative concentration of lysine was 1.54-times greater in the high calprotectin group (95% CI 1.05-2.03, P = 0.028).

Conclusions: This multiomic study identified an increase in *Dialister* species and pentanoate, and a decrease in lysine, in patients with "mucosal healing." It supports further investigation of these as potential novel therapeutic targets in CD.

The Handling Editor for this article was Professor Jonathan Rhodes, and it was accepted for publication after full peer-review.

1 | INTRODUCTION

Crohn's disease (CD) is a chronic inflammatory disease caused by an inappropriate immunological response to the gastrointestinal microbiota in genetically susceptible individuals. Transmural inflammation throughout the gastrointestinal tract causes pain, bleeding, weight loss and diarrhoea; ultimately resulting in long-term complications such as stricture and fistula formation. Children with CD typically have more aggressive disease and worse long-term outcomes than those with adult onset disease.[1]

Treatment of paediatric CD aims to control symptoms and to suppress mucosal inflammation. Adequate suppression of inflammation results in mucosal healing, which is associated with improved long-term outcomes.[1,2] Low faecal calprotectin levels are a well validated non-invasive marker of mucosal healing in CD.[3-6] The faecal calprotectin threshold which defines mucosal healing has not been universally agreed, studies support values in the range of 100-300 μg/g,[3-8] with a value of 100 μg/g used most frequently.[9] In paediatric CD Weinstein-Nakar et al[7] found faecal calprotectin levels under 100 μg/g were predictive of mucosal and transmural healing.

Inducing and maintaining mucosal healing is a major challenge. Over 25% of all paediatric CD patients are now escalated to anti-TNF and other biological therapies,[10] but approximately one in three patients will not have a sustained response to maximal medical therapy.[11,12]

The gastrointestinal microbiome has an important role in mucosal immune homoeostasis and evidence from epidemiological studies and mouse models suggest it is involved in the pathogenesis of CD.[13,14] There are several mechanisms by which the microbiota could promote mucosal healing or exacerbate mucosal inflammation in established disease.[15]

Changes to the microbiota's metabolic capacity alter the mucosal concentration of active metabolites. The short-chain fatty acids (SCFAs) acetate, propionate, butyrate and pentanoate (also known as valerate) are produced by bacterial metabolism of dietary fibre and amino acids,[16] and are present in reduced concentration in CD.[17] SCFAs act as a primary energy source for enterocytes, promote tight junction formation and induce anti-microbial protein formation.[18] Butyrate[18,19] and pentanoate[20] induce anti-inflammatory cytokine production and modify immune cell differentiation via their activation of free fatty acid receptors and inhibition of histone deacetylase. Immunomodulatory properties of other metabolites, including bile acid metabolites[21] and amino acids,[22] have also been described.

Protease activity in the gastrointestinal lumen is influenced by microbial metabolism. Bacteria produce proteases, and also degrade or inhibit host pancreatic and inflammatory proteases. Protease activity is increased in IBD,[23] and can influence disease activity by inducing invasive behaviour in commensal bacteria[24] or by degrading tight junction proteins.[25] Pro-inflammatory signalling cascades are initiated by protease activated receptors on enterocytes, fibroblasts and mucosal immune cells.[24,25]

Previous studies have characterised the microbiome of paediatric CD patients relative to controls and found reduced diversity, high instability and altered relative abundance of multiple bacterial genera.[26,27] Increased mucosal inflammation has been associated with a greater reduction in alpha diversity, as well as reduced abundance of SCFA producers such as *Faecalibacterium prausnitzii*[26] and *Roseburia hominis*.[28,29] Conversely, inducing remission with anti-TNF induction makes the microbiota more similar to that of healthy children.[30] It is not known whether inflammation associated microbial differences persist in established disease, or whether they contribute to ongoing mucosal healing or inflammation.

This study aims to identify changes in the microbiota which are associated with mucosal healing in established paediatric CD and to investigate the functional implications of these changes using inferred metagenomics, ^{1}H-NMR metabolomics and protease activity assays.

2 | MATERIALS AND METHODS

2.1 | Study design and participant recruitment

Participants were recruited to the "Metabolic and Microbiome Profiling in Paediatric IBD" study, a multi-centre clinical cohort study in North-West London. The parents or carers of all study participants provided informed consent to join the study, and the study participants signed age appropriate assent forms. The study design and objectives were approved by London Bloomsbury research ethics committee (REC reference 17/LO/2049). All patients included in this analysis had a confirmed diagnosis of paediatric CD and a disease duration of greater than 6 months. All were receiving maintenance biological therapy, which was defined as regular treatment with infliximab, adalimumab or vedolizumab beyond the induction phase.

Clinical data were obtained from review of electronic medical records and from a questionnaire completed at induction to the study. C-reactive protein, erythrocyte sedimentation rate, haemoglobin concentration and stool calprotectin concentration were tested as part of routine clinical care.

All participants provided a sample of a whole stool for microbiome, metabolome and protease activity analysis. Stool samples were collected using a "fecotainer" system (Excretas Medical). The sample was immediately refrigerated to 4°C and was homogenised, divided and frozen at minus 80°C within 4 hours of production. Sample order was randomised prior to further analysis.

The low calprotectin group was defined by a calprotectin concentration of less than 100 μg/g, which was used as a surrogate for mucosal healing, and the high calprotectin group by a concentration of 100 μg/g or over. Calprotectin value was used to assign groups if it was measured in the same stool sample as that used for the study analyses, or if two stable calprotectin values were recorded within the 2 months before and after collection of the sample used for study analyses and there had been no change in clinical disease activity or treatment in this period. Samples without a corresponding calprotectin measurement meeting these requirements were not included in the high and low calprotectin groups, but were included for analyses which did not require

calprotectin value, that is, correlation analyses between microbiome and metabolome.

Baseline characteristics were compared between groups using Welch t test for continuous variables, and the chi-squared test for categorical variables (where test validity requirements were met).

2.2 | Metataxonomic analysis

Metataxonomic analysis (16S rRNA gene sequencing) was performed on DNA extracted from thawed stool samples. The V1-2 region of the 16S rRNA gene was used for taxonomic discrimination.[31-33]

2.2.1 | DNA extraction and sequencing

DNA was extracted using the Powerlyzer PowerSoil DNA isolation kit (Qiagen) according to the manufacturers protocol except samples were lysed by bead beating for 3 minutes at speed 8 using a Bullet Blender Storm instrument (Thistle Scientific). Sample libraries were prepared using a protocol adapted from Illumina's 16S Metagenomic Sequencing Library Preparation Protocol,[34] which has been previously described.[35] The V1-V2 region of 16S rRNA gene were amplified using the primers listed in Table S1. SequalPrep Normalization Plate Kit (Thermo Fischer Scientific) was used to normalise index PCR reactions, and NEBNext Library Quant Kit for Illumina (New England Biolabs) was used to quantify sample libraries. Sequencing was performed on an Illumina MiSeq platform (Illumina Inc) using paired-end 300 bp chemistry and the MiSeq Reagent Kit v3 (Illumina).

2.3 | Data analysis

16S rRNA gene sequence data were analysed using the DADA2 pipeline (version 1.12)[36] in R (R Foundation). Outlier analysis (ROUT, $Q = 1$) of the number of sequences from each sample was performed using GraphPad Prism (GraphPad Software). The vegan package in R was used to calculate Shannon index, species richness and Pielou evenness for all samples. A neighbour-joining phylogenetic tree was constructed using the R packages DECIPHER and phangorn, and the phyloseq package[37] was used to calculate weighted UniFrac distances. Weighted UniFrac distance was compared between groups using non-metric multidimensional scaling (NMDS) and permutational analysis of variance (PERMANOVA), performed using the R packages vegan and ggplot 2.

The amplicon sequence variant (ASV) table produced by DADA2 was formatted by grouping rare variants followed by subsampling data to the level of the sample with the lowest summed read count which was 7114 reads. No samples were identified as an outlier for summed read count using GraphPad Prism (ROUT, $Q = 1$). STAMP[38] was used to analyse difference in relative abundance between high and low calprotectin groups at all taxonomic levels from genus to phylum. Two group comparison was performed in STAMP using two-sided White's non-parametric t test with Benjamini-Hochberg correction for multiple testing ($q < 0.05$ was considered significant). Results were filtered to select only those where the difference in the mean proportion of sequences was greater than 1%.

2.3.1 | Inferred metagenomic analysis

Inferred metagenomic analysis was performed using Piphillin, which is the best performing tool for predicting metagenome composition in clinical biosamples.[39] Three outputs were analysed: KEGG pathways, KEGG orthologs and BioCyc features. All were formatted by subsampling to the level of the sample with the lowest summed counts (1.51×10^6, 7.98×10^6, and 2.84×10^6 respectively). Two-group comparison was performed in STAMP using the same parameters as above. An effect size filter of 0.1% difference between pathway, ortholog or feature expression was used.

2.4 | Protease activity assay

A mixture of one-part thawed stool sample and two parts phosphate-buffered saline (PBS) were vortexed with microbeads at 2850 revolutions per minute for 5 minutes using the Vortex Genie 2 with microcentrifuge tube adaptor (Scientific industries Inc). The homogenised mixture was centrifuged at 20 000 ***g*** for 20 minutes. Faecal water was prepared by passing 500 µl supernatant through 5.0 µm centrifugal filter (12 000 ***g***, 3 minute spin). PBS was added in appropriate volume to prepare 10- and 100-fold dilutions of faecal water.

Protein concentration of faecal water dilutions was measured and compared against bovine serum albumin standards using the colorimetric Pierce BCA Protein assay following the manufacturer's protocol (Thermo Fischer Scientific). Protein concentration in the faecal water was normalised to 1mg/ml by addition of PBS. Dilutions of the 1mg/ml faecal water and trypsin standard were prepared. These were added to an FTC-Casein substrate and working reagent from the Pierce Fluorescent Protease Assay kit (Thermo Fischer Scientific) as per the manufacturers protocol. Proteolysis of the substrate was measured by fluorescent resonance energy transfer using a FLUOstar OPTIMA (BMG LABTECH). Total protease activity of faecal water was expressed relative to the activity of trypsin standards in ng/ml concentration (relative units). All assays performed in triplicate.

The activity of protease subtypes was assessed by performing the above protease activity assay, but with the co-incubation of a protease inhibitor (Roche Diagnostics) with the faecal solution for 5 minutes, prior to addition of the working reagent. Optimal inhibitor concentrations were determined in preliminary work and inhibitor specificity confirmed by searching the MEROPS database.[40] The following inhibitors were used: Antipain-dihydrochloride (50 µg/ml), Leupeptin (5 µg/ml), or a combination of antipain-dihydrochloride

(50 μg/ml), aprotinin (2 μg/ml) and Pefabloc (1 mg/ml). Assays were performed in duplicate. Degree of inhibition was calculated as ratio of protease activity in inhibited and uninhibited wells. Protease subtype activity (Δ protease activity) was calculated as total protease activity multiplied by degree of inhibition. Outliers for protease and protease subtype activity were identified using GraphPad Prism (ROUT, Q = 1) and excluded from further analysis. Welch's t test was performed to compare total protease and protease subtype activity between groups. Bonferroni correction was used to correct p values from protease subtype activity analysis.

2.5 | Metabolome analysis

2.5.1 | Sample preparation for NMR analysis

Faecal water extraction was carried out by homogenising each sample with water (UHPLC grade, Fisher Chemical) at a ratio of 1:2 (mg of wet weight of faecal sample: μL of water). The mixture was vortexed for 5 minutes and centrifuged at 18 000 ***g*** and 4°C for 20 minutes, 540 μL of the supernatant were mixed with 60 μL of 1.5 M KH_2PO_4 buffer (pH 7.4, 100% of deuterium oxide (D_2O), 2 mM sodium azide and 1% of TSP (3-trimethylsilyl-[2,2,3,3,-$^2H_4$]-propionic acid sodium salt). The mixture was centrifuged at 18 000 ***g*** at 4°C for 1 min. An aliquot of 580 μL of the supernatant was transferred into a 5 mm outer diameter NMR tube as described previously.[41]

2.5.2 | ^{1}H-NMR spectroscopy

Water-suppressed ^{1}H NMR spectroscopy was performed at 300 K on a Bruker 600 MHz Avance III HD spectrometer (Bruker Biospin). The 1D ^{1}H NMR spectra were acquired using standard one-dimensional pulse sequence, with saturation of the water resonance (noesygppr1d pulse program) during both the relaxation delay (RD = 4s) and mixing time (t_m = 10 ms). The two magnetic field z-gradients implemented by this pulse sequence are applied for 1 ms The receiver gain was set to 90.5 and acquisition time (ACQ) to 2.73s for all experiments. Each 1-dimensional (1D) ^{1}H-NMR spectrum was acquired using 4 dummy scans, 32 scans, 64 K data points and with a spectral window of 20 ppm. Prior to Fourier Transformation, each free induction decay was multiplied by an exponential function corresponding to a line broadening of 0.3 Hz.

2D ^{1}H – ^{1}H *J*-resolved experiment was also acquired for each sample to detect the *J*-couplings in the second dimension using the pulse program with suppression of the water resonance during the relaxation delay (jresgpprqf). The acquisition parameters used for this experiment were as follows: 16 dummy scans and 2 scans, 8K points with spectral window of 16.7 ppm for f2 and 40 increments with a spectral window of 78 Hz for f1, incremented delay of 3μs, RD of 2 s and ACQ of 0.41 s. The receiver gain was set to 90.5. A sine-bell apodization function was applied on both dimensions, followed by Fourier transformation, tilting by 45°, and symmetrisation along f1.[42]

2D-NMR experiments including ^{1}H–^{1}H Total Correlation Spectroscopy (TOCSY) and ^{1}H–^{13}C Heteronuclear Single Quantum Coherence (HSQC) were acquired for a representative sample for identification purposes.

2.5.3 | Spectral data analysis

1D ^{1}H-NMR spectra were phased, baseline corrected and calibrated to TSP peak at δ^1H 0 ppm prior to being digitised over the range δ^1H −0.5 to 11 ppm, and imported into MATLAB (2014a, MathWorks). Spectral regions containing residual water (δ^1H 4.69-4.93), TSP (δ^1H −0.50 to 0.62) and noise (δ^1H 9.35-11.00) were removed prior to probabilistic quotient normalisation.[43]

The data set was auto-scaled and modelled using Partial Least Squares Discriminant Analysis (PLS-DA) in a Monte Carlo Cross-Validation (MCCV) framework using Storey-Tibshirani method of correction for multiple testing.[44] Variables with $q < 0.05$ were considered to be significant. Goodness of fit (R^2Y) was calculated using the training data, and the goodness of prediction (Q^2Y) from test data. For discriminant analysis, calprotectin was used as dummy variable (*Y* = 0 and 1 for high and low calprotectin respectively).

2.5.4 | Identification of metabolites

The structural identification of significant metabolites was achieved by 2D-NMR experiments and Subset Optimization by Reference Matching (STORM) on 1D ^{1}H-NMR data set.[45] Internal and external databases such as the Human Metabolome Data Base (HMDB; http://hmdb.ca/)[46] and/or the Biological Magnetic Resonance Data Bank (BMRB; http://www.bmrb.wisc.edu) were used for confirmation of assignments.

Relative concentrations were calculated for metabolites that had significant correlation with Dialister and those that were observed as discriminators on the MCCV-PLS-DA model. This was performed by taking the integral from normalised spectra of a representative resonance peak of pentanoate at 1.30 (δ^1H range from 1.27 to 1.33), lysine at 1.73 (δ^1H range from 1.69 to 1.76), beta-alanine at 2.56 (δ^1H range from 2.55 to 2.58) and phenylalanine at 7.43 ppm (δ^1H range from 7.41 to 7.45). The integral of a peak represents a relative concentration of its corresponding metabolite. Mean metabolite relative concentration was compared between high and low calprotectin groups.

2.5.5 | Metataxonomic-Metabolome correlation analysis

Full resolution NMR spectral features were correlated with relative abundance data at five taxonomic levels (genera – phyla) using Spearman's rank correlation coefficient with Storey-Tibshirani method for multiple testing correction. A threshold of $|\rho| \geq 0.5$ with $q < 0.05$ was used to indicate a significant correlation.

2.5.6 | Metabolic network analysis

Metabolites that had significant correlation with Dialister and those that were observed as discriminators on the MCCV-PLS-DA model were used to construct a condensed multicompartmental metabolic reaction network using MetaboNetworks v.2.3.[47] This software is an agnostic bioinformatic tool that can reconstruct the pathway relationship observed in a real system, which is visualised as classical representation (KEGG). It calculates the shortest paths (number of reactions) between metabolites, including only the reactions that can occur in humans considered as a supra-organism. A custom metabolic reaction network was constructed using enzymes linked to *Homo sapiens* genes and enzymes linked to bacteria present in the gastrointestinal tract. Species from genera with >1.5% relative abundance in either the high or low calprotectin group (Figure S2), and that had genomic data available on the KEGG organism database were selected for inclusion in the analysis.

2.5.7 | Metabolic potential of Dialister species

In order to identify the most appropriate reference genome to assess the metabolic potential of *Dialister*, species level analysis was performed for the *Dialister* genus. The 334 base pair ASVs from *Dialister* (Table S2) were matched against 16S sequences in the rRNA/ITS database (National Centre for Biotechnology Information (NCBI), updated 4 July 2020) using the megaBLAST[48] nucleotide search tool. Species identity was assigned if there was 98.7% sequence match to a reference strain.[49] This confidence threshold has high accuracy for species identification when using the V1-2 region of the 16S rRNA gene.[33]

The NCBI reference sequence linked to the representative strain was used to generate a metabolic pathway map using MetaCyc (https://metacyc.org).[50]

2.6 | Sensitivity analyses

The metataxonomic, protease activity and metabolome data analysis were repeated for two sensitivity analyses. The first was performed due to the uncertainty surrounding the appropriate threshold of faecal calprotectin which indicates mucosal healing. Values of 100-300 μg/g have been suggested as appropriate thresholds. Participants with faecal calprotectin concentrations between 100 and 300 μg/g were therefore excluded, leaving only those with faecal calprotectin levels which are widely accepted to predict mucosal healing or mucosal inflammation.

The second sensitivity analysis excluded participants who were receiving vedolizumab. Vedolizumab is an $\alpha_4\beta_7$ integrin antagonist, it has a different mechanism of action compared to anti-TNF agents and could conceivably induce distinct changes in the microbiota. Participants receiving vedolizumab were excluded in order to remove this as a potential confounding factor.

3 | RESULTS

3.1 | Participant characteristics

The "Metabolic and Microbiome Profiling in Paediatric IBD" study recruited 53 children. Of these, 25 patients with established paediatric CD on maintenance biological therapy were eligible for inclusion in this analysis (Figure S1). There were 11 patients in each of the low and high calprotectin groups. Three patients met inclusion criteria, but did not have a calprotectin value, these patients were included in analyses which did not require calprotectin value.

Height was significantly lower in the high calprotectin group (difference in mean *z* score −1.22, 95% CI −2.26 to −0.18, $P = 0.025$). There were no significant differences in other baseline characteristics. Disease location, and treatment with biological and immunomodulatory therapy were similar between groups (Table 1).

Clinical disease activity was not significantly different between groups. Mean paediatric CD activity index (PCDAI) ± standard deviation was 19 ± 22 in the high calprotectin group vs 13.8 ± 14 in the low calprotectin group ($P = 0.50$). Median calprotectin in the low calprotectin group was 38 μg/g (range 23-97) compared to 1338 μg/g (range 141-3933) in the high calprotectin group ($P = 0.0016$).

3.2 | Metataxonomic analysis

Bacterial diversity metrics and community composition were similar in the high and low calprotectin groups. There was no significant difference in Shannon diversity index, Pielou's evenness index or species richness between groups (Figure 1A). No samples were identified as outliers by these diversity metrics. An NMDS plot of weighted UniFrac distance showed no clear separation of the two groups by beta diversity (Figure 1B). This finding was consistent with non-significance on PERMANOVA of weighted UniFrac distance ($P = 0.33$). The sample with the highest calprotectin concentration (3933 μg/g) was identified as an outlier on weighted UniFrac distance and was not included in the NMDS plot or PERMANOVA calculation.

There was a sixfold increase in the relative abundance of the *Dialister* genus in the low calprotectin group compared to the high calprotectin group. *Dialister* species were the seven most abundance genus in the low calprotectin group with a mean proportion of sequences of 3.4% (SD 2.22), compared to the 30 most abundant in the high calprotectin group with a mean proportion of sequences of 0.53% (SD 0.79) (difference of means −2.88%, 95% CI −4.31 to −1.53, $q = 0.00999$; Figure 2). The relative abundance of Dialister was negatively correlated with calprotectin (Spearman's $\rho = -0.73$, $P = 0.0001$). There were no other significant differences in relative abundance with $q < 0.05$ and difference in the mean proportion of sequences >1% at genus level (Figure S2), or at higher taxonomic levels.

Inferred metagenomic analysis identified no significant differences in KEGG ortholog, KEGG pathway or BioCyc feature expression between the two groups.

TABLE 1 Baseline characteristics of the low and high calprotectin groups

	Low calprotectin	High calprotectin	*P* value
Number	11	11	
Age, mean ± SD (y)	13.4 ± 3.30	14.4 ± 2.20	0.414
Disease duration, mean ± SD (mo)	30.7 ± 13.5	29.8 ± 13.8	0.878
Time on biologics, Mean ± SD (mo)	18.6 ± 11.4	19.1 ± 13.8	0.921
Height (*z* score)	0.534 ± 1.35	−0.689 ± 0.689	0.025
BMI (*z* score)	0.174 ± 1.17	0.0648 ± 1.24	0.846
Biological therapy	Infliximab 91%	Infliximab 91%	
	Vedolizumab 9%	Vedolizumab 9%	0.99
Immunomodulator	Azathioprine 18%	Azathioprine 45%	—
	Methotrexate 18%	Methotrexate 0%	
	Mercaptopurine 9%	Mercaptopurine 0%	
Disease location	L1 0%	L1 9%	—
	L2 45%	L2 36%	
	L3 45%	L3 45%	
	L4a/b 54%	L4a/b 63%	
	Perianal 9%	Perianal 27%	

Note: P values calculated using unpaired unequal variance t test (Welch's test) for continuous variables, and chi-squared test for categorical variables where validity requirements met.

3.3 | Protease activity

Protease activity was assayed in all thawed stool samples. Outlier analysis detected an anomalously high protease activity of 1809 relative units in the sample with the highest calprotectin concentration (3933 μg/g), this result was removed from further analysis of protease and protease subtype activity.

There was no significant difference in total protease activity between the groups. Mean protease activity was 587 relative units (SD 227) in the high calprotectin group vs 465 relative units (SD 254) in the low calprotectin group (difference of means 121, 95% CI −99 to 342, $P = 0.26$; Figure 3A).

The activity of protease subtypes was assayed by addition of different combinations of protease inhibitors to control the spectrum of inhibition. After Bonferroni correction there was no significant difference in the activity of any protease subtypes (Figure 3B-D).

3.4 | ^{1}H-NMR-based metabolic profiling

^{1}H-NMR spectroscopy was performed on faecal water prepared from thawed stool samples. Analysable spectra were not obtained for three samples from the high calprotectin group and one sample without a calprotectin value. The remaining data set was modelled using high and low calprotectin groups as a dichotomous classifier (n = 19). The MCCV-PLS-DA model discriminated samples from high and low calprotectin groups with explained variance (R^2Y) of 0.87 and capability of prediction (Q^2Y) of 0.41 (Figure 4A).

The principal metabolic discriminators used by the model were pentanoate, which was increased in the low calprotectin group, and lysine which was increased in the high calprotectin group (Figure 4). Peak assignments of both metabolites were confirmed by multiple two-dimensional NMR experiments and STORM method (Table S3 and Figure S4). Other regions of the 1D spectra detected as significant by the model did not correspond to assignable metabolites (Figure 4B).

The relative concentrations of pentanoate and lysine in the high and low calprotectin groups were approximated using the integrals of their resonance peaks at 1.30 ppm (δ^1H range from 1.27 to 1.33) and 1.73 (δ^1H range from 1.69 to 1.76) respectively (Figure 4C). The mean concentration of pentanoate was 1.35 times greater in the low calprotectin group compared to the high calprotectin group (95% CI 1.03-1.68 times greater, $P = 0.0361$) and had strong negative correlation with calprotectin ($\rho = -0.61$, $P = 0.006$). The relative concentration of lysine was 1.54 times greater in the high calprotectin group (95% CI 1.05-2.03 times greater, $P = 0.028$) and had strong positive correlation with calprotectin ($\rho = 0.58$, $P = 0.009$).

3.5 | Multiomic integration

Relative abundance data from metataxonomic analysis were correlated with 1D ^{1}H-NMR spectral features using Spearman's rank correlation coefficient with Storey-Tibshirani correction for multiple testing (n = 21). *Dialister* species abundance had significant

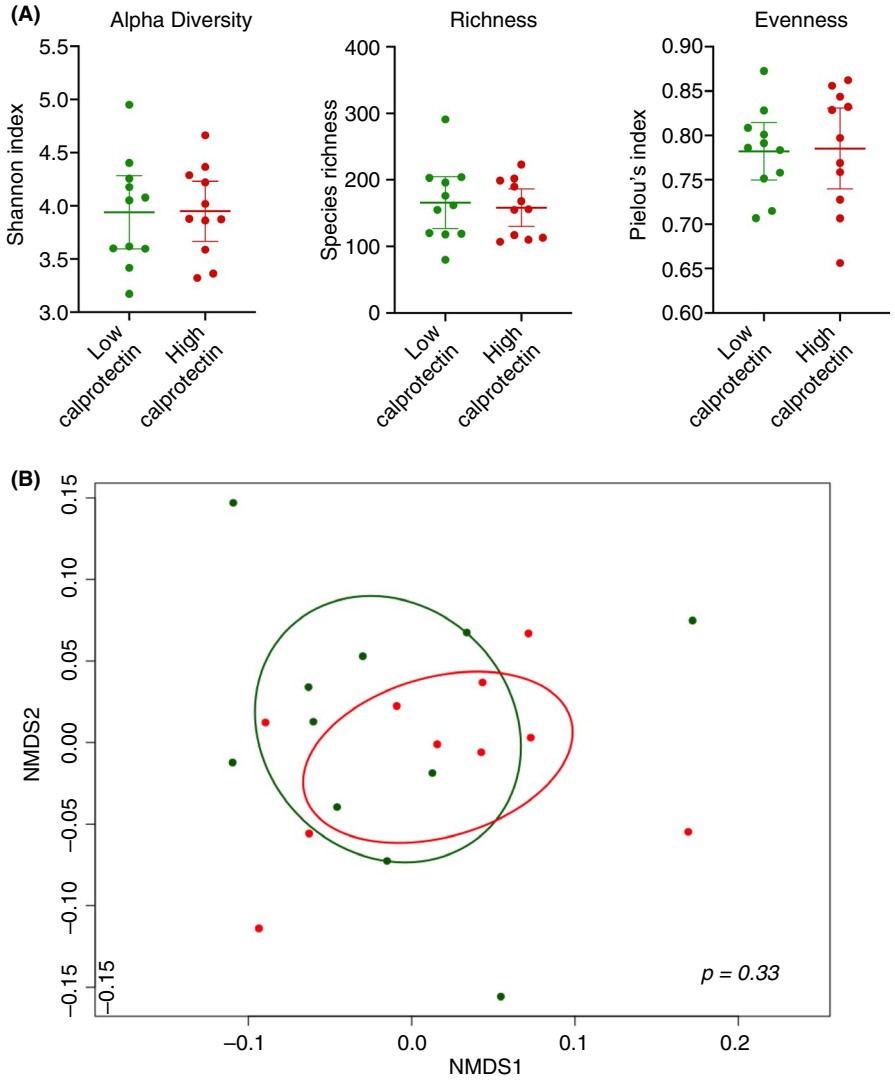

FIGURE 1 Comparison of microbial diversity metrics and community composition between high and low calprotectin groups. A, Scatter plots with mean and 95% confidence interval for alpha diversity (Shannon index), species richness, and evenness (Pielou's index). B, Non-metric multidimensional scaling (NMDS) plot of weighted UniFrac distance with 95% confidence ellipses. Red dots and ellipse = high calprotectin, green dots and ellipse = low calprotectin. *P* value calculated using PERMANOVA. [Correction added on September 21, 2020, after first online publication: A typo in the word "calprotectin" in the X-axes of Figure 1A has been corrected.]

positive correlation ($\rho > 0.5$, $P < 0.05$) with the concentration of pentanoate and ß-alanine, and significant negative correlation ($\rho < -0.5$, $P < 0.05$) with phenylalanine (Figure S3 and Table S4). The presence of these metabolites was confirmed by multiple two-dimensional NMR approaches and STORM method (Table S4 and Figure S4). Unlike pentanoate, there was no significant difference in the relative concentration of beta-alanine or phenylalanine between the high and low calprotectin groups ($P = 0.25$ and $P = 0.12$ respectively). Four other bacterial genera (*Escherichia/Shigella*, *Unclassified* Lachnospiraceae, *Roseburia* and *Veillonella*) had significant correlations with metabolites, as shown in Table S4.

3.6 | Metabolic potential of *Dialister* species

Metataxonomic and metabolite data were used to create a condensed reaction network in MetaboNetworks (Figure S5).[47] This illustrates metabolic pathways available in the microbiome of participants, and how they lead to biosynthesis or degradation of significant metabolites.

The potential contribution of *Dialister* to the production or degradation of significant metabolites was assessed by comparing its metabolic network map to the microbiome reaction network. Species level analysis of the *Dialister* genus was performed to identify the most appropriate reference genome to use to generate its metabolic network map. This revealed that 79% of the difference in the *Dialister* genus relative abundance between the groups was attributable to *D invisus*, and that the relative abundance of *D invisus* was significantly higher in the low calprotectin group ($P = 0.0069$) (Table S2). A metabolic network map was created from the genome of a representative strain from the *D invisus* species (*D invisus* JCM 17566). This network map acts as a partial approximation of the true metabolic potential of *Dialister* species, it is unable to account for variation among the different strains of the species, and does not contain all metabolic processes which occur within this strain.

The metabolic network of *D invisus* JCM 17566 includes the succinate pathway, which is a major route for propionate formation. Propionate can be coupled to yield pentanoate (Figure S5). The sequenced genome of *D invisus* JCM 17566 does not feature any annotated genes for the biosynthesis or degradation of beta-alanine or phenylalanine.

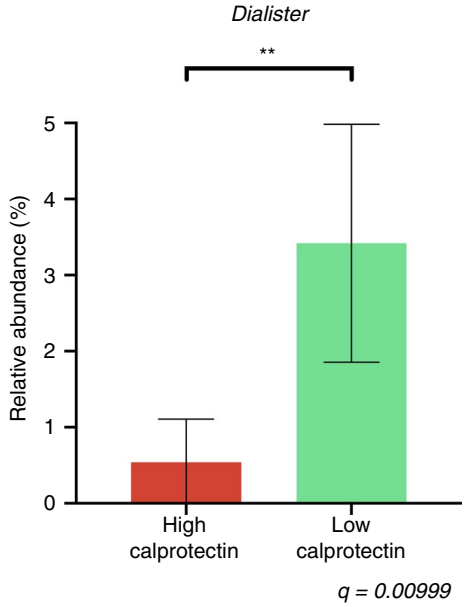

FIGURE 2 Mean relative abundance of *Dialister* in high and low calprotectin groups, bars indicate 95% confidence intervals. *q* value calculated using two-sided White's non-parametric t test with Benjamini-Hochberg correction for multiple testing. [Correction added on September 21, 2020, after first online publication: A typo in the word "calprotectin" in the X-axis of Figure 2 has been corrected.]

3.7 | Sensitivity analyses

Two sensitivity analyses were performed. The first excluded participants with faecal calprotectin values within the range of suggested thresholds for mucosal healing (100-300 μg/g), leaving only those with calprotectin values very highly predictive of mucosal healing or inflammation. One patient from the high calprotectin group with a faecal calprotectin value of 141 μg/g was excluded. The second sensitivity analysis excluded patients receiving maintenance therapy with vedolizumab rather than an anti-TNF agent. This removed potential bias resulting from different therapy induced changes to the microbiota. Two patients were excluded, one from each group. The metataxonomic and metabolomic findings of the primary analysis were reproduced in both sensitivity analyses (Table S5).

4 | DISCUSSION

This study investigates changes in the microbiota associated with mucosal healing, as indicated by calprotectin level <100 μg/g, in a cohort of patients with established paediatric CD receiving maintenance biological therapy. There were no significant differences in alpha or beta diversity between the high and low calprotectin groups.

Interestingly, there was significantly greater relative abundance of bacteria from the *Dialister* genus in patients who had low calprotectin levels, and the relative abundance of *Dialister* had a strong negative correlation with calprotectin. The difference in *Dialister* relative abundance was largely attributable to the *D invisus* species.

Dialister is obligately anaerobic or microaerophilic members of the firmicutes phylum. Among healthy adults they have been noted to have a bimodal abundance distribution, being either abundant or nearly absent, with decreased temporal stability in intermediate states.[51] *Dialister* species are known to have reduced abundance in adult and paediatric CD patients relative to healthy controls.[52,53] Interestingly *Dialister* also has higher relative abundance in CD patients who are in remission following ileocaecal resection than those who have post-operative recurrence.[54] It is possible that *Dialister*, with its bimodal abundance distribution, acts as a biomarker of mucosal inflammation. Alternatively, *Dialister* may influence disease activity via its metabolic activity.

An estimate of *D invisus's* metabolic potential was made by producing a metabolic map from its protein coding genes which have been linked to metabolic pathways. This is not able to provide a complete account of its metabolic activity but did identify potential production of propionate. Production of propionate and acetate by *D invisus* has been confirmed in culture experiments.[55] Propionate is utilised in pentanoate formation. The positive correlation of *Dialister* with pentanoate observed in our study is consistent with *Dialister* promoting the formation of pentanoate, either through increasing the availability of its precursor or by a more direct mechanism.

Pentanoate, also known as valerate, was identified as a key indicator of mucosal healing by a machine-learning model utilising spectral data from ^{1}H-NMR metabolomic analysis to distinguish high and low calprotectin groups. The relative concentration of pentanoate was higher in the low calprotectin group and was negatively correlated with calprotectin. Pentanoate has previously been found to be deficient in stool samples from patients with CD compared to controls,[17] and an inverse relationship with clinical disease activity has been described in adult CD.[56,57] In paediatric CD levels of pentanoate have been shown to increase with successful exclusive enteral nutrition treatment.[58] An association with mucosal healing has not previously been described.

In vitro evidence shows that pentanoate has marked immunomodulatory effects on lymphocytes through promoting AKT-mTOR signalling and enhancing histone acetyltransferase activity, leading to increased IL-10 and suppressed IL-17 production.[20] Pentanoate treatment decreased Th17 cell differentiation and reduced acute inflammation in a mouse model of experimental autoimmune encephalitis.[20] In addition to its immunomodulatory effects pentanoate has also been found to impair vegetative growth of *Clostridioides difficile* in vitro, and supplementation has been proposed as a possible treatment strategy in *C difficile* infection.[59] The association of pentanoate with mucosal healing demonstrated in our study is consistent with its described effects on lymphocytes and pathogenic microbiota. It supports a model in which reduced concentration of pentanoate contributes to the pathogenesis of CD.

Lysine was also identified as a discriminator by the machine learning model; its relative concentration was higher in the high calprotectin group and was positively correlated with calprotectin. Faecal lysine has previously been found to be present in higher concentration in CD than in healthy controls,[60] but it is not known whether it has any biological significance. Lysine is known to competitively inhibit uptake of arginine at the CAT1 transporter, and thus could reduce uptake of arginine which has immunomodulatory properties.[61] Alternatively, the increase in lysine may be secondary to inflammation induced changes in microbial amino acid metabolism, and could be a useful biomarker of disease activity.

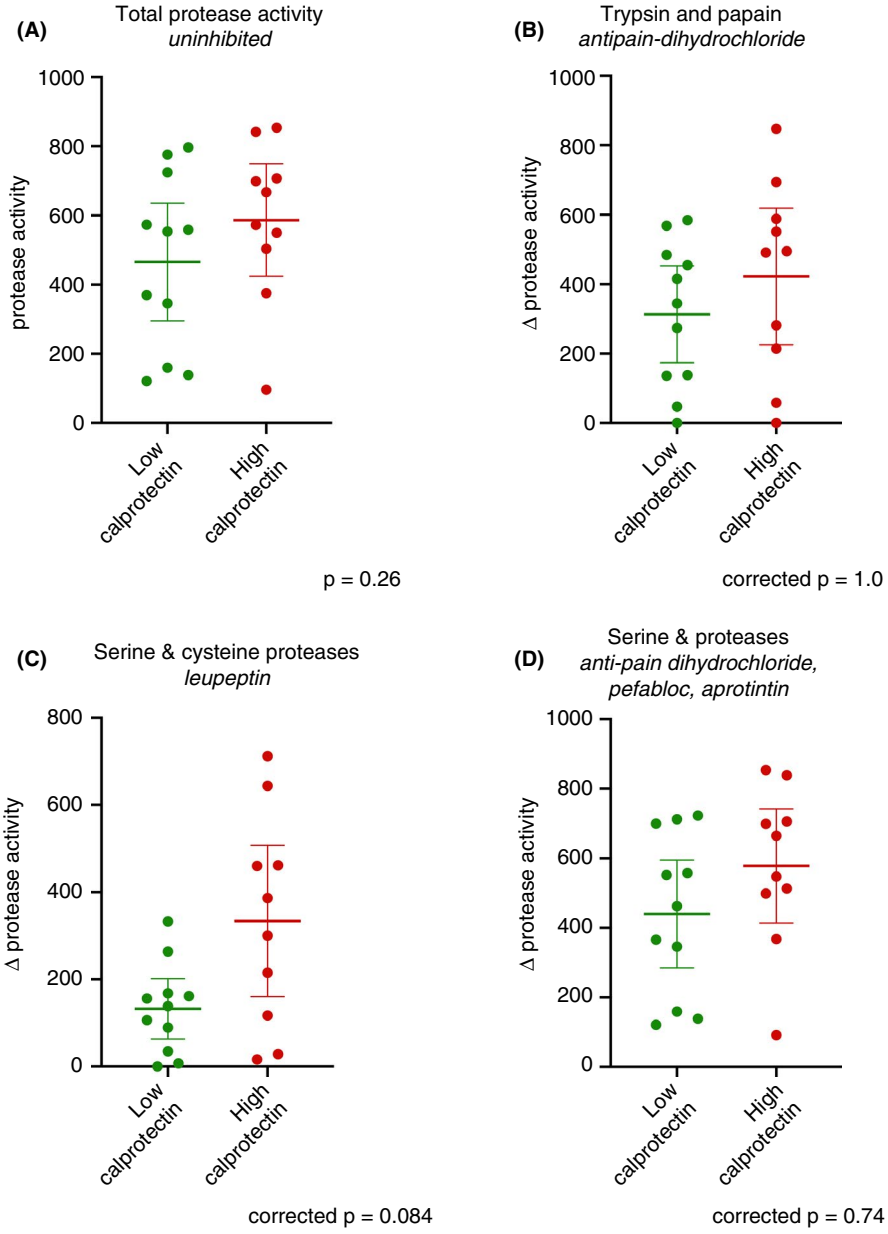

FIGURE 3 Scatter plots of protease activity (expressed as equivalent concentration of trypsin in μg/ml) in patients with high and low calprotectin. Bars indicate mean and 95% confidence interval. Unpaired t test was performed for all samples, Bonferroni correction was applied for protease subtype activity. A, Total protease activity (uninhibited). B-D, Protease subtype activity, expressed as reduction in protease activity caused by addition of inhibitors (Δ protease activity); antipain-dihydrochloride (B); leupeptin (C); antipain-dihydrochloride, pepstatin, aprotinin (D). [Correction added on September 21, 2020, after first online publication: A typo in the word "calprotectin" in the X-axes of Figure 3 has been corrected.]

Beta-alanine and phenylalanine were correlated with *Dialister*, however, their relative concentrations were not significantly different between the high and low calprotectin groups and the machine learning model did not identify either metabolite as a discriminator. It is not clear that beta-alanine or phenylalanine are associated with mucosal healing.

Given the significant metataxonomic and metabolomic differences between the high and low calprotectin groups it is surprising no differences were detected by inferred metagenomic analysis. However, apart from *Dialister* there were no large-scale changes in microbial community composition in our study, and the lack of difference likely reflects the limitations of inferred approaches in detecting smaller metagenomic differences.

No difference in total protease activity or protease subtype activity was found between high and low calprotectin groups. Protease subtype analysis was limited by non-specificity of protease inhibitors, therefore several significant protease classes such as matrix-metalloproteases could not be assessed.

This study has several limitations. It is a cross sectional analysis of 25 participants, longitudinal analysis of a larger data set would give further insights on how the microbiota and metabolome change with disease activity and would aid detection of smaller changes. Use of metagenomic sequencing would enhance understanding of how changes in the microbiota influence the metabolome, and targeted metabolomic approaches would allow quantification and improved detection of metabolites.

Use of a faecal calprotectin level under 100 μg/g as a biomarker for mucosal healing is a potential limitation of this study, as there is uncertainty regarding the appropriate threshold to indicate mucosal healing, with values in the range of 100-300 μg/g supported. A sensitivity analysis which excluded patients with "borderline" faecal calprotectin values was performed to address this and confirmed all

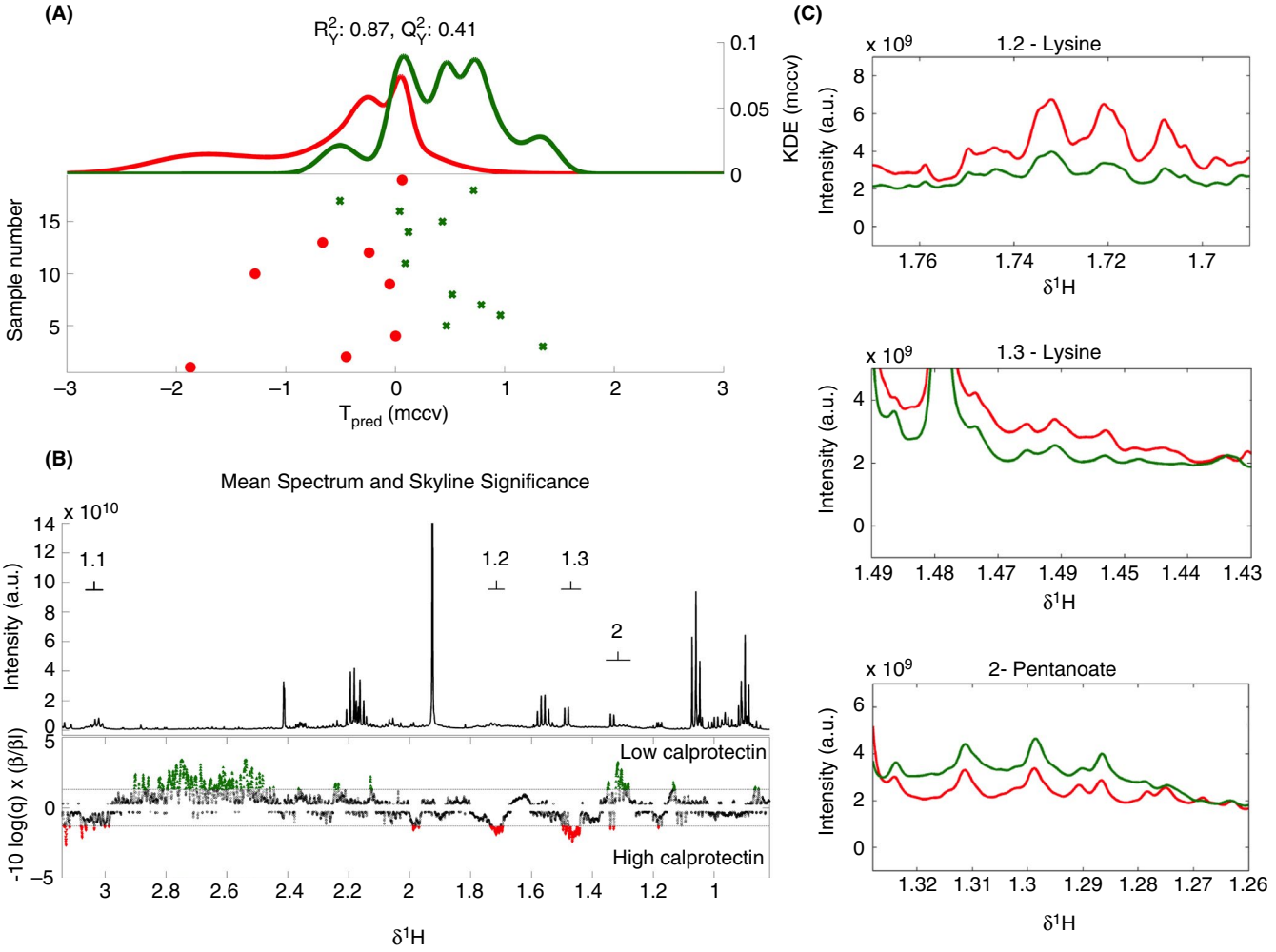

FIGURE 4 A, MCCV-PLS-DA score plot derived from 1D ^{1}H-NMR spectra of faecal water samples, indicating the differentiation between high calprotectin group (red) and low calprotectin group (green). The model is comprised of Kernel Density Estimate (KDE) of the predicted scores (Tpred) for both groups. Dots represent the metabolic profile of each patient from the study cohort when its corresponding NMR spectrum and calprotectin values were available (n = 19). The fit and predictability of the model were obtained and expressed as R^2Y (explained variance) and Q^2Y (capability of prediction) values. B, MCCV-PLS-DA loading plot. Top depicts the average 1D ^{1}H-NMR spectrum of the 19 faecal water samples. The bottom depicts the Manhattan plot showing $-\log_{10}(q) \times$ sign of regression coefficient (β) of the MCCV-PLS-DA model that in conjunction represent the contribution of each variable on it. In green, metabolites are shown that are significantly higher in the low calprotectin group, and in red metabolites significantly higher in the high calprotectin group. Labels: 1.1-1.3 = lysine (1.73(m), 1.46(m), 3.04 (t)), 2 = pentanoate/valerate (1.30(m)). C, Fragments from the average 1D ^{1}H-NMR spectra from the high calprotectin (red, n = 8) and low calprotectin (green, n = 11) groups, showing the less overlapped resonances detected as significant in the MCCV-PLS-DA model corresponding to lysine (1.2, 1.73(m), 1.3, 1.46(m)) and pentanoate/valerate (2, 1.30(m)). MCCV, Monte Carlo Cross-Validation; PLS-DA, partial least squares discriminant analysis

findings from the primary analysis. There is additional concern the accuracy of faecal calprotectin may be lower in ileal disease,[9] but in this cohort only one patient had isolated ileal disease and they had a very raised calprotectin level over 2000 μg/g.

In this study clinical disease activity did not correlate closely with calprotectin. This observation is consistent with many other studies in adult and paediatric CD and highlights the need to consider both clinical remission and mucosal healing as distinct therapeutic targets.[62]

This study addresses an important and clinically relevant question—are there microbial or metabolic features in established paediatric CD which could contribute to sustained mucosal inflammation or promote mucosal healing? It reports a comprehensive multiomic analysis of a well-phenotyped cohort, which is representative of many patients with established paediatric CD. It has several important findings. Lysine is increased in patients with ongoing mucosal inflammation, this association warrants further investigation to identify potentially relevant disease mechanisms and biomarkers. Most significantly, *D invisus* is more abundant in patients with mucosal healing and has the potential to facilitate pentanoate formation. Pentanoate, which has known immunomodulatory properties highly relevant to CD, is also increased in those with mucosal healing and is correlated with *Dialister* relative abundance. The *Dialister* pentanoate relationship could therefore represent a novel and targetable therapeutic axis which promotes mucosal healing in CD.

ACKNOWLEDGEMENTS

We thank Dr Matthew Hyde and Professor Modi of the Neonatal research unit at Chelsea and Westminster Hospital for providing clinical research facilities, and Dr Warren Hyer, Eunice Goto and Kay Crook for facilitating recruitment and sample collection. We also thank all of the patients and their families for their participation.

Declaration of personal interests: Ailsa Hart has served as consultant, advisory board member or speaker for AbbVie, Arena, Atlantic, Bristol-Myers Squibb, Celgene, Celltrion, Falk, Ferring, Janssen, MSD, Napp Pharmaceuticals, Pfizer, Pharmacosmos, Shire and Takeda. She also serves on the Global Steering Committee for Genentech. Jia Li is supported by Medical Research Council (MRC) New Investigator Research Grant (MR/P002536/1) and ERC Starting Grant (715662). Dr JM Fell has served as an advisory board member for Falk Pharma. Henry Taylor, Jose Ivan Serrano Contreras, Rocio Seoane, Julie McDonald, Jenny Epstein, and Julian Marchesi have no personal or funding interests to declare.

AUTHORSHIP

Guarantor of the article: Ailsa Hart.

Author contributions: HT, JE, JF, JRM and AH were involved in study conception and design. HT was involved in participant recruitment, sample and data collection and first draft of manuscript. HT, JISC, JAKM and RCS were involved in sample analysis. HT, JISC, JAKM, JVL and JRM were involved in data analysis. All authors approved the final manuscript.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Henry Taylor https://orcid.org/0000-0001-9066-9151
Jose Ivan Serrano-Contreras https://orcid.org/0000-0001-8669-7571
Julie A. K. McDonald https://orcid.org/0000-0003-0739-6047
Jenny Epstein https://orcid.org/0000-0003-1380-465X
Rocio C. Seoane https://orcid.org/0000-0002-2938-3872
Jia V. Li https://orcid.org/0000-0002-5763-6670

REFERENCES

1. Ruemmele FM, Veres G, Kolho KL, et al. Consensus guidelines of ECCO/ESPGHAN on the medical management of pediatric Crohn's disease. *J Crohns Colitis*. 2014;8:1179-1207.
2. Pineton de Chambrun G, Peyrin-Biroulet L, Lémann M, Colombel J-F. Clinical implications of mucosal healing for the management of IBD. *Nat Rev Gastroenterol Hepatol*. 2010;7:15-29.
3. Vazquez Moron JM, Pallares Manrique H, Machancoses FH, Ramos Lora M, Ruiz FC. Accurate cut-offs for predicting endoscopic activity and mucosal healing in Crohn's disease with fecal calprotectin. *Rev Esp Enferm Dig*. 2017;109:130-136.
4. Mosli MH, Zou G, Garg SK, et al. C-reactive protein, fecal calprotectin, and stool lactoferrin for detection of endoscopic activity in symptomatic inflammatory bowel disease patients: a systematic review and meta-analysis. *Am J Gastroenterol*. 2015;110(6):802-819; quiz 820.
5. Kennedy NA, Jones G-R, Plevris N, Patenden R, Arnott ID, Lees CW. Association between level of fecal calprotectin and progression of Crohn's disease. *Clin Gastroenterol Hepatol*. 2019;17:2269-2276.e4.
6. Kostas A, Siakavellas SI, Kosmidis C, et al. Fecal calprotectin measurement is a marker of short-term clinical outcome and presence of mucosal healing in patients with inflammatory bowel disease. *World J Gastroenterol*. 2017;23:7387-7396.
7. Weinstein-Nakar I, Focht G, Church P, et al. Associations among mucosal and transmural healing and fecal level of calprotectin in children with Crohn's disease. *Clin Gastroenterol Hepatol*. 2018;16:1089-1097.
8. Reenaers C, Bossuyt P, Hindryckx P, Vanpoucke H, Cremer A, Baert F. Expert opinion for use of faecal calprotectin in diagnosis and monitoring of inflammatory bowel disease in daily clinical practice. *United Eur Gastroenterol J*. 2018;6:1117-1125.
9. Simon EG, Wardle R, Thi AA, Eldridge J, Samuel S, Moran GW. Does fecal calprotectin equally and accurately measure disease activity in small bowel and large bowel Crohn's disease? A systematic review. *Intest Res*. 2019;17:160-170.
10. Ashton JJ, Borca F, Mossotto E, et al. Increased prevalence of anti-TNF therapy in paediatric inflammatory bowel disease is associated with a decline in surgical resections during childhood. *Aliment Pharmacol Ther*. 2019;49:398-407.
11. Ding NS, Hart A, De Cruz P. Systematic review: predicting and optimising response to anti-TNF therapy in Crohn's disease—algorithm for practical management. *Aliment Pharmacol Ther*. 2016;43:30-51.
12. Grossi V, Lerer T, Griffiths A, et al. Concomitant use of immunomodulators affects the durability of infliximab therapy in children with Crohn's disease. *Clin Gastroenterol Hepatol*. 2015;13:1748-1756.
13. Shouval DS, Rufo PA. The role of environmental factors in the pathogenesis of inflammatory bowel diseases: a review. *JAMA Pediatr*. 2017;171:999-1005.
14. Gkouskou KK, Deligianni C, Tsatsanis C, Eliopoulos AG. The gut microbiota in mouse models of inflammatory bowel disease. *Front Cell Infect Microbiol*. 2014;4:28.
15. Lopetuso LR, Petito V, Zambrano D, et al. Gut microbiota: a key modulator of intestinal healing in inflammatory bowel disease. *Dig Dis*. 2016;34:202-209.
16. Neis EPJG, Dejong CHC, Rensen SS. The role of microbial amino acid metabolism in host metabolism. *Nutrients*. 2015;7:2930-2946.
17. Zhuang X, Li T, Li M, et al. Systematic review and meta-analysis: short-chain fatty acid characterization in patients with inflammatory bowel disease. *Inflamm Bowel Dis*. 2019;25:1751-1763.
18. Parada Venegas D, De la Fuente MK, Landskron G, et al. Short chain fatty acids (SCFAs)-Mediated gut epithelial and immune regulation and its relevance for inflammatory bowel diseases. *Front Immunol*. 2019;10:277.
19. Schulthess J, Pandey S, Capitani M, et al. The short chain fatty acid butyrate imprints an antimicrobial program in macrophages. *Immunity*. 2019;50(2):432-445.e7.
20. Luu M, Pautz S, Kohl V, et al. The short-chain fatty acid pentanoate suppresses autoimmunity by modulating the metabolic-epigenetic crosstalk in lymphocytes. *Nat Commun*. 2019;10:760.
21. Duboc H, Rajca S, Rainteau D, et al. Connecting dysbiosis, bile-acid dysmetabolism and gut inflammation in inflammatory bowel diseases. *Gut*. 2013;62:531-539.
22. Liu Y, Wang X, Hu C-AA. Therapeutic potential of amino acids in inflammatory bowel disease. *Nutrients*. 2017;9:920. https://doi.org/10.3390/nu9090920
23. Jablaoui A, Kriaa A, Mkaouar H, et al. Fecal serine protease profiling in inflammatory bowel diseases. *Front Cell Infect Microbiol*. 2020;10:21.
24. Carroll IM, Maharshak N. Enteric bacterial proteases in inflammatory bowel disease- pathophysiology and clinical implications. *World J Gastroenterol*. 2013;19:7531-7543.
25. Vergnolle N. Protease inhibition as new therapeutic strategy for GI diseases. *Gut*. 2016;65:1215-1224.
26. Gevers D, Kugathasan S, Denson L, et al. The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe*. 2014;15:382-392.

27. Goyal A, Yeh A, Bush BR, et al. Safety, clinical response, and microbiome findings following fecal microbiota transplant in children with inflammatory bowel disease. *Inflamm Bowel Dis*. 2018;24:410-421.
28. Kolho K-L, Korpela K, Jaakkola T, et al. Fecal microbiota in pediatric inflammatory bowel disease and its relation to inflammation. *Am J Gastroenterol*. 2015;110:921-930.
29. Lewis J, Chen E, Baldassano R, et al. Inflammation, antibiotics, and diet as environmental stressors of the gut microbiome in pediatric Crohn's Disease. *Cell Host Microbe*. 2015;18:489-500.
30. Kowalska-Duplaga K, Kapusta P, Gosiewski T, et al. Changes in the intestinal microbiota are seen following treatment with infliximab in children with Crohn's disease. *J Clin Med*. 2020;9:687.
31. Gołębiewski M, Tretyn A. Generating amplicon reads for microbial community assessment with next-generation sequencing. *J Appl Microbiol*. 2020;128:330-354.
32. Alcon-Giner C, Caim S, Mitra S, et al. Optimisation of 16S rRNA gut microbiota profiling of extremely low birth weight infants. *BMC Genom*. 2017;18:841.
33. Johnson JS, Spakowicz DJ, Hong B-Y, et al. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat Commun*. 2019;10:5029.
34. Illumina. 16S metagenomic sequencing library preparation [Online]. https://support.illumina.com/downloads/16s_metagenomic_sequencing_library_preparation.html. Accessed March 26, 2019.
35. Mullish BH, Pechlivanis A, Barker GF, Thursz MR, Marchesi JR, McDonald JAK. Functional microbiomics: evaluation of gut microbiota-bile acid metabolism interactions in health and disease. *Methods*. 2018;149:49-58.
36. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods*. 2016;13:581-583.
37. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*. 2013;8:e61217.
38. Parks DH, Tyson GW, Hugenholtz P, Beiko RG. STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics*. 2014;30:3123-3124.
39. Iwai S, Weinmaier T, Schmidt BL, et al. Piphillin: improved prediction of metagenomic content by direct inference from human microbiomes. *PLoS One*. 2016;11:e0166104.
40. Rawlings ND, Waller M, Barrett AJ, Bateman A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res*. 2014;42(Database issue):D503–9.
41. Gratton J, Phetcharaburanin J, Mullish BH, et al. Optimized sample handling strategy for metabolic profiling of human feces. *Anal Chem*. 2016;88:4661-4668.
42. Dona AC, Jiménez B, Schäfer H, et al. Precision high-throughput proton NMR spectroscopy of human urine, serum, and plasma for large-scale metabolic phenotyping. *Anal Chem*. 2014;86:9887-9894.
43. Dieterle F, Ross A, Schlotterbeck G, Senn H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. *Anal Chem*. 2006;78:4281-4290.
44. Posma JM, Garcia-Perez I, Ebbels TMD, et al. Optimized phenotypic biomarker discovery and confounder elimination via covariate-adjusted projection to latent structures from metabolic spectroscopy data. *J Proteome Res*. 2018;17:1586-1595.
45. Posma JM, Garcia-Perez I, De Iorio M, et al. Subset optimization by reference matching (STORM): an optimized statistical approach for recovery of metabolic biomarker structural information from 1H NMR spectra of biofluids. *Anal Chem*. 2012;84:10694-10701.
46. Wishart DS, Feunang YD, Marcu A, et al. HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res*. 2018;46(D1):D608-D617.
47. Posma JM, Robinette SL, Holmes E, Nicholson JK. MetaboNetworks, an interactive Matlab-based toolbox for creating, customizing and exploring sub-networks from KEGG. *Bioinformatics*. 2013;30:893-895.
48. Morgulis A, Coulouris G, Raytselis Y, Madden TL, Agarwala R, Schäffer AA. Database indexing for production MegaBLAST searches. *Bioinformatics*. 2008;24:1757-1764.
49. Yarza P, Yilmaz P, Pruesse E, et al. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol*. 2014;12:635-645.
50. Caspi R, Billington R, Ferrer L, et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res*. 2016;44:D471-D480.
51. Lahti L, Salojärvi J, Salonen A, Scheffer M, de Vos WM. Tipping elements in the human intestinal ecosystem. *Nat Commun*. 2014;5:4344.
52. Joossens M, Huys G, Cnockaert M, et al. Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. *Gut*. 2011;60:631-637.
53. Kowalska-Duplaga K, Gosiewski T, Kapusta P, et al. Differences in the intestinal microbiome of healthy children and patients with newly diagnosed Crohn's disease. *Sci Rep*. 2019;9:18880.
54. Mondot S, Lepage P, Seksik P, et al. Structural robustness of the gut mucosal microbiota is associated with Crohn's disease remission after surgery. *Gut*. 2016;65:954-962.
55. Sakamoto M, Ikeyama N, Toyoda A, et al. *Dialister hominis* sp. nov., isolated from human faeces. *Int J Syst Evol Microbiol*. 2020;70:589-595.
56. De Preter V, Machiels K, Joossens M, et al. Faecal metabolite profiling identifies medium-chain fatty acids as discriminating compounds in IBD. *Gut*. 2015;64:447-458.
57. De Preter V, Joossens M, Ballet V, et al. Metabolic profiling of the impact of oligofructose-enriched inulin in Crohn's disease patients: a double-blinded randomized controlled trial. *Clin Transl Gastroenterol*. 2013;4:e30.
58. Tjellström B, Högberg L, Stenhammar L, et al. Effect of exclusive enteral nutrition on gut microflora function in children with Crohn's disease. *Scand J Gastroenterol*. 2012;47:1454-1459.
59. McDonald JAK, Mullish BH, Pechlivanis A, et al. Inhibiting growth of *Clostridioides difficile* by restoring valerate, produced by the intestinal microbiota. *Gastroenterology*. 2018;155:1495-1507.e15.
60. Marchesi JR, Holmes E, Khan F, et al. Rapid and noninvasive metabonomic characterization of inflammatory bowel disease. *J Proteome Res*. 2007;6:546-551.
61. Sugihara K, Morhardt TL, Kamada N. The role of dietary nutrients in inflammatory bowel disease. *Front Immunol*. 2019;9:3183.
62. Peyrin-Biroulet L, Sandborn W, Sands BE, et al. Selecting therapeutic targets in inflammatory bowel disease (STRIDE): determining therapeutic goals for treat-to-target. *Am J Gastroenterol*. 2015;110:1324-1338.

SUPPORTING INFORMATION

Additional supporting information will be found online in the Supporting Information section.

How to cite this article: Taylor H, Serrano-Contreras JI, McDonald JAK, et al. Multiomic features associated with mucosal healing and inflammation in paediatric Crohn's disease. *Aliment Pharmacol Ther*. 2020;00:1–12. https://doi.org/10.1111/apt.16086